

(3) Japanese Patent Application Laid-Open No. 8-221090 (1996) "SPEECH RECOGNITION METHOD"

The following is English translation of an extract from the above-identified document relevant to the present application.

This invention is related to a speech recognition method, and in particular, related to a speech recognition method of frame synchronized processing which uses a Viterbi algorithm by using Hidden Markov Model.

A network by Hidden Markov Model is expressed by using states and node. On this network, according to a Viterbi algorithm, all of the items required for recognition processing are combined with cumulative collation score and then propagated and processed for each speech recognition candidate produced in each state.

(19)日本国特許庁 (JP)

(12)公開特許公報 (A)

(11)特許出願公開番号

特開平8-221090

(43)公開日 平成8年(1996)8月30日

(51)Int.Cl.
G10L 3/00

識別記号
535

F I
G10L 3/00
535

審査請求 未請求 請求項の数11 O L (全8頁)

(21)出願番号

特願平7-26379

(22)出願日

平成7年(1995)2月15日

(71)出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72)発明者 實廣 貴敏

東京都千代田区内幸町1丁目1番6号 日本電信電話株式会社内

(72)発明者 嵐峨山 茂樹

東京都千代田区内幸町1丁目1番6号 日本電信電話株式会社内

(74)代理人 弁理士 草野 順

(54)【発明の名称】音声認識方法

(57)【要約】

【目的】 累積照合スコアの計算量を減らし、記憶量も比較的小さく、隠れマルコフモデルにまたがった継続時間長制御をする音声認識方法を提供する。

【構成】 隠れマルコフモデル自身を構成する確率分布に対応づけられている状態と認識処理の途中情報を保持して入力の終端において最終的な累積照合スコアがネットワークのどこを通過してきたかを得る節点とより成る隠れマルコフモデルのネットワークにより認識候補を表現し、これら状態系列を統合し、認識処理時に隠れマルコフモデルの照合スコアを各状態で計算し、計算された累積照合スコアを隣接する状態または節点に伝播していく時に隠れマルコフモデルと組み合わせて他の情報をまとめて伝播させ、各状態または節点からの遷移の時に認識処理を進める音声認識方法。

【特許請求の範囲】

【請求項1】 入力信号を特徴パラメータの系列に変換し、変換された特徴パラメータ系列について、1個もしくは複数個の認識候補に対して、隠れマルコフモデルと特徴パラメータ系列とをそれぞれ分析フレーム単位で照合処理して、Viterbi アルゴリズムにより尤度の高い候補を認識結果とする音声認識方法において、隠れマルコフモデル自身を構成する確率分布に対応づけられている状態と認識処理の途中情報を保持して入力の終端において最終的な累積照合スコアがネットワークのどこを通過してきたかを得る節点とより成る隠れマルコフモデルのネットワークにより認識候補を表現し、これら状態系列を統合し、認識処理時に隠れマルコフモデルの照合スコアを各状態で計算し、計算された累積照合スコアを隣接する状態または節点に伝播して行く時に隠れマルコフモデルと組み合わせて他の情報をまとめて伝播させ、各状態または節点からの遷移の時に認識処理を進めることを特徴とする音声認識方法。

【請求項2】 請求項1に記載される音声認識方法において、各状態または節点は前々時刻の累積照合スコアと前時刻の出現確率との間の積を保持し、現時刻において遷移の計算時に遷移確率との間の積をとって、前時刻の累積照合スコアを計算し、遷移していく経路の中で最大の累積照合スコアを残しておき、現時刻に対する出現確率を計算し、累積照合スコアとその出現確率との間の積を保持して、次時刻に使用する操作を繰り返すことにより累積照合スコアの計算をすることを特徴とする音声認識方法。

【請求項3】 請求項1に記載される音声認識方法において、各隠れマルコフモデルを構成する状態の系列より長い状態系列でネットワークを構成し、認識処理に必要な項目を伝播させ、この長い状態系列に対する制御することを特徴とする音声認識方法。

【請求項4】 請求項3に記載される音声認識方法において、状態以上の長さが音素、音節、単語、或は文節であることを特徴とする音声認識方法。

【請求項5】 請求項4に記載される音声認識方法において、他の情報が状態遷移が生じた時刻である場合に、これを累積照合スコアと共に伝播し、継続時間長制御を行うことを特徴とする音声認識方法。

【請求項6】 請求項4に記載される音声認識方法において、他の情報がその状態に達するまでの音素、音節、単語、或は文節の言語的履歴である場合に、これを累積照合スコアと共に伝播し、言語的履歴の連鎖確率を言語的制約として利用することを特徴とする音声認識方法。

【請求項7】 請求項4に記載される音声認識方法において、他の情報がその状態に達するまでの継続時間長の履歴である場合に、これを累積照合スコアと共に伝播し、継続時間長の履歴より現在の継続時間長を予測して継続時間長制御することを特徴とする音声認識方法。

【請求項8】 請求項4に記載される音声認識方法において、他の情報を状態遷移が生じた時刻およびその状態に達するまでの音素、音節、単語、或は文節の言語的履歴としてこれらの情報を累積照合スコアと共に伝播し、継続時間長制御を行うと共に言語的履歴の連鎖確率を言語的制約として利用することを特徴とする音声認識方法。

【請求項9】 請求項4に記載される音声認識方法において、他の情報を状態遷移が生じた時刻およびその状態に達するまでの継続時間長の履歴としてこれらの情報を累積照合スコアと共に伝播し、継続時間長制御をすると共に継続時間長の履歴より現在の継続時間長を予測して継続時間長制御することを特徴とする音声認識方法。

【請求項10】 請求項4に記載される音声認識方法において、他の情報をその状態に達するまでの音素、音節、単語、或は文節の言語的履歴およびその状態に達するまでの継続時間長の履歴としてこれら的情報を累積照合スコアと共に伝播し、言語的履歴の連鎖確率を言語的制約として利用すると共に継続時間長の履歴より現在の継続時間長を予測して継続時間長制御することを特徴とする音声認識方法。

【請求項11】 請求項4に記載される音声認識方法において、他の情報を状態遷移が生じた時刻、その状態に達するまでの音素、音節、単語、或は文節の言語的履歴およびその状態に達するまでの継続時間長の履歴としてこれらの情報を累積照合スコアと共に伝播し、継続時間長制御をすると共に言語的履歴の連鎖確率を言語的制約として利用し、更に継続時間長の履歴より現在の継続時間長を予測して継続時間長制御することを特徴とする音声認識方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 この発明は、音声認識方法に関し、特に、隠れマルコフモデルを使用し、Viterbi アルゴリズムを使用するフレーム同期で処理する音声認識方法に関する。

【0002】

【従来の技術】 従来、隠れマルコフモデルと組み合わせて他の情報をも使用して音声認識候補を表現して音声認識をする場合、音声認識候補を木構造により広げて行く仕方を実施してきた（例えば、言語モデルを用いる場合、T.Kawabata, T.Hanazawa, K.Itoh and K.Shikano: "Japanese Phonetic Typewriter Using HMM Phone Recognition and Stochastic Phone-Sequence Modeling," IEICE Trans., E74(7), pp.1783-1787(1991)）。しかし、この仕方は、認識候補を隠れマルコフモデルのネットワークにより表現し、これら認識候補の状態系列を統合することができるネットワークにおいて音声認識途中で実施することはできない。また、木構造は、累積照合スコアの計算が枝の数だけ多くなる問題がある。

【0003】 更に、隠れマルコフモデルは一つの状態に

滞留する時間の情報が明確に反映されるものではない。そのために、隠れマルコフモデルに基づいたフレーム同期でViterbiアルゴリズムを使用する音声認識方法においては、語彙の制約がない場合に挿入誤りが多く出現する。そこで、従来は、隠れマルコフモデルの時間的構造を補うために、隠れマルコフモデルを構成する状態毎に、或はモデル単位である音素毎に継続時間長制御を行っていた。

【0004】

【発明が解決しようとする課題】この発明は、隠れマルコフモデルによるネットワークを状態と節点により表現し、このネットワーク上においてViterbiアルゴリズムにより各状態に生じる音声認識候補について認識処理に必要な項目をすべて累積照合スコアと組にして伝播、処理することにより、累積照合スコアの計算量を減らし、記憶量も比較的小さく、隠れマルコフモデルにまたがった継続時間長制御をすることができる音声認識方法を提供するものである。

【0005】

【課題を解決するための手段】入力信号を特徴パラメータの系列に変換し、変換された特徴パラメータ系列について、1個もしくは複数個の認識候補に対して、隠れマルコフモデルと特徴パラメータ系列とをそれぞれ分析フレーム単位で照合処理して、Viterbiアルゴリズムにより尤度の高い候補を認識結果とする音声認識方法において、隠れマルコフモデル自身を構成する確率分布に対応づけられている状態と認識処理の途中情報を保持して入力の終端において最終的な累積照合スコアがネットワークのどこを通過してきたかを得る節点とより成る隠れマルコフモデルのネットワークにより認識候補を表現し、これら状態系列を統合し、認識処理時に隠れマルコフモデルの照合スコアを各状態で計算し、計算された累積照合スコアを隣接する状態または節点に伝播して行く時に隠れマルコフモデルと組み合わせて他の情報をまとめて伝播させ、各状態または節点からの遷移の時に認識処理を進める音声認識方法を構成した。

【0006】そして、各状態または節点は前々時刻の累積照合スコアと前時刻の出現確率との間の積を保持し、現時刻において遷移の計算時に遷移確率との間の積をとつて、前時刻の累積照合スコアを計算し、遷移していく経路の中で最大の累積照合スコアを残しておき、現時刻に対する出現確率を計算し、累積照合スコアとその出現確率との間の積を保持して、次時刻に使用する操作を繰り返すことにより累積照合スコアの計算をする音声認識方法を構成した。

【0007】また、各隠れマルコフモデルを構成する状態の系列より長い状態系列でネットワークを構成し、認識処理に必要な項目を伝播させ、この長い状態系列に対する制御をする音声認識方法を構成した。更に、状態以上の長さが音素、音節、単語、或は文節である音声認識

方法を構成した。

【0008】ここで、他の情報が状態遷移が生じた時刻である場合に、これを累積照合スコアと共に伝播し、継続時間長制御を行う音声認識方法を構成した。また、他の情報がその状態に達するまでの音素、音節、単語、或は文節の言語的履歴である場合に、これを累積照合スコアと共に伝播し、言語的履歴の連鎖確率を言語的制約として利用する音声認識方法を構成した。更に、他の情報がその状態に達するまでの継続時間長の履歴である場合に、これを累積照合スコアと共に伝播し、継続時間長の履歴より現在の継続時間長を予測して継続時間長制御をする音声認識方法を構成した。

【0009】また、他の情報を状態遷移が生じた時刻およびその状態に達するまでの音素、音節、単語、或は文節の言語的履歴としてこれらの情報を累積照合スコアと共に伝播し、継続時間長制御を行うと共に言語的履歴の連鎖確率を言語的制約として利用する音声認識方法を構成した。更に、他の情報を状態遷移が生じた時刻およびその状態に達するまでの継続時間長の履歴としてこれらの情報を累積照合スコアと共に伝播し、継続時間長制御をすると共に継続時間長の履歴より現在の継続時間長を予測して継続時間長制御をする音声認識方法を構成した。

【0010】そして、他の情報をその状態に達するまでの音素、音節、単語、或は文節の言語的履歴およびその状態に達するまでの継続時間長の履歴としてこれらの情報を累積照合スコアと共に伝播し、言語的履歴の連鎖確率を言語的制約として利用すると共に継続時間長の履歴より現在の継続時間長を予測して継続時間長制御をする音声認識方法を構成した。

【0011】また、他の情報を状態遷移が生じた時刻、その状態に達するまでの音素、音節、単語、或は文節の言語的履歴およびその状態に達するまでの継続時間長の履歴としてこれらの情報を累積照合スコアと共に伝播し、継続時間長制御をすると共に言語的履歴の連鎖確率を言語的制約として利用し、更に継続時間長の履歴より現在の継続時間長を予測して継続時間長制御をする音声認識方法を構成した。

【0012】

【実施例】この発明は、隠れマルコフモデル（「確率モデルによる音声認識」、中川 聖一、電子情報通信学会編（1988）参照）を使用し、Viterbiアルゴリズムを使用したフレーム同期で処理する音声認識方法に関する。そして、ここに言う音声は、楽器音、超音波その他の音響信号に代表される時系列、或は、手書き文字を構成する線素に代表される空間的特徴量系列、その他の特徴パラメータ系列一般を含むものとする。

【0013】この発明の第1の実施例を図1を参照して説明する。図1はこの発明における隠れマルコフモデルのネットワークの一例を示す。図1に示される例は、

「平等」「病院」「病人」「描写」の4個の単語から成るネットワークである。単語を対応する音素系列に変換し、これら音素系列を更に対応する隠れマルコフモデルに変換し、隠れマルコフモデルの状態系列のネットワークを構成する。図1における白丸は隠れマルコフモデル自身を構成する状態を表し、黒丸はその状態系列を統合する節点である。ここで、接点は、音声終端で最大スコアを持つ経路を後でたどるために、各時刻(フレーム)で選択された経路の接続されている方向およびその節点に到達する前に別の節点を通った時の時刻を各時刻に対して保持しておく。このために、接点を通過する時刻を累積照合スコアと同時に伝播させる必要がある。図1の様に隠れマルコフモデルのネットワークを構成すれば、例示される4個の単語間の共通する内容の部分「びよー」は統合して累積照合スコアの大きいものだけを計算すれがよく、計算量を減少することができる。そして、

$$L_i(t+1) = \max \{ L_i(t)b_i(x_i)a_{i1}, L_i(t)b_i(x_i)a_{i2} \} \quad (1)$$

但し、 $L_i(t)$ 、 $L_i(t)$ ：時刻(分析フレーム) t における累積照合スコア

$b_i(x_i)$ 、 $b_i(x_i)$ ：出現確率

a_{i1} 、 a_{i2} ：遷移確率

この計算を行う場合、遷移してきた累積照合スコア $L_i(t+1)$ を計算するのに、度々、前の状態における出現確

$$L'_i(t+1) = \max \{ L'_i(t)a_{i1}, L'_i(t)a_{i2} \} \quad (2)$$

$$L'_i(t+1) = \max \{ L'_i(t)a_{i1}, L'_i(t)a_{i2} \} b_i(t+1) \quad (3)$$

即ち、図3に示される様に、 $L'_i(t+1)$ を各状態に保持することにすれば、遷移してきた累積照合スコアの大きい方に、その状態の出現確率を乗算すればよいことになる。これにより、各状態の出現確率計算の度毎に前の状態を見に行く必要はなくなる。但し、 $t=0$ のとき L'_i

$$L_i(T) = L'_{i1}a_{i11111} \quad (4)$$

この様に、効率的な累積照合スコアの計算を実現することができる。この発明の第2の実施例を図4を参照して説明する。第2の実施例は、隠れマルコフモデルによる標準パターンの他に情報として状態遷移の生じた時刻をも使用し、継続時間長制御を行う例である。累積照合スコアと継続時間長制御を行なう際に使用される状態遷移時刻とを組にして伝播する。白丸は隠れマルコフモデルの状態を表し、この状態を時間的に並べたものについて認識処理の説明をする。状態間を結合する太い矢印はViterbiアルゴリズムにより生じる1つの経路を示す。破線の間に継続時間長制御の単位、例えば音節、単語に対応している。各時刻、各状態について隠れマルコフモデルの累積照合スコアを持ち、更に継続時間長単位で遷移してきた時刻を保持し、これを伝播していくと考える。先ず、時刻 t_1 に前の継続時間長の単位から累積照合スコアが遷移してきたものとする。この時の累積照合スコア $P(t_1)$ と共に遷移時刻 t_1 を記憶する。次の時刻 t_2 の時、この経路は同じ状態に留まり、累積照合スコアは $P(t_1)$ に更新されるが、時刻はそのまま t_1 を保持す

る。これと節点より成るネットワークを構成することにより、何れの状態も前の状態または節点からのリンクを1本だけ有すればよく、記憶量の比較的少ないネットワークを構成することができる。また、各累積照合スコアの遷移に対応して、各節点にのみ時刻を残しておくことにより、音声の終端において認識結果を得ることができる。

【0014】Viterbiアルゴリズムは、状態遷移の際、遷移してきた累積照合スコアと自己ループの累積照合スコアとを比較し、値の大きい方を残すアルゴリズムである。各フレーム毎にこの処理を実施し、音声の終端においては、各節点に残した時刻と接続情報を元に、最終的なネットワーク上の最大スコアを持つ経路を得ることができ、この経路に対応する単語、文その他の出力用に対応付けたものを認識結果とする。図2は隠れマルコフモデルの一部の定義を示す図であるが、この定義から、

率 $b_i(x_i)$ 、 $b_i(x_i)$ を得る必要がある。このオーバーハンドは、各状態に保持する値を工夫することにより回避することができる。

【0015】ここで、 $L'_i(t)=L_i(t)b_i(x_i)$ とおくと、

$$L'_i(t+1) = \max \{ L'_i(t)a_{i1}, L'_i(t)a_{i2} \} \quad (2)$$

$$L'_i(t+1) = \max \{ L'_i(t)a_{i1}, L'_i(t)a_{i2} \} b_i(t+1) \quad (3)$$

$(0)=1$ 、 $L'_i(0)=0$ ($j \neq 0$)であり、また、 $t=T$ 、終了時刻のとき、終端の節点Nについて遷移してくる累積尤度を計算する必要がある。

【0016】

$$L_i(T) = L'_{i1}a_{i11111} \quad (4)$$

る。これを継続時間長単位の最終状態まで続ける。図4の経路は、時刻 t_1 において継続時間長単位間の遷移が生じるが、ここで、先ず、継続時間長 $\gamma=t_2-t_1$ を求める。この γ に対して継続時間長の出現確率を求め、これを累積照合スコア $P(t_1)$ に反映させる。この時、経路に保持する時刻を t_1 に更新し、以後、同様に伝播していく。この様にして累積照合スコアと共に遷移時刻を組にして伝播することにより、状態より長い系列を単位とする継続時間長制御をすることができる。

【0017】図5は、この発明による音声認識装置のブロック図である。1は入力端子、2は特徴抽出部である。特徴抽出部2は、入力端子1から入力される音声をデジタル信号に変換し、更に、これを1フレーム例えば10ms毎に特徴パラメータに変換してその特徴量を出力する。この特徴パラメータとしては、例えば、LP Cケプストラム係数を採用することができる。

【0018】3はネットワーク探索部である。このネットワーク探索部3においては、各フレームにおいて、先ず、累積照合スコア計算部4から累積照合スコアを得

る。次いで、Viterbiアルゴリズムに基づいて、幾つかの経路が交わる状態或は節点において累積照合スコアが最大になる経路のみを残す。節点においては、選択された経路を後でたどることができる様に、前の節点からの時刻とその節点に対する接続方向の情報を残す。これらを音声の終端に達するまでフレーム毎に繰り返し、音声終端においてはネットワーク上の最大累積照合スコアを持つ経路を選択して認識結果を得る。

【0019】6は音素モデルであり、学習用音声データベースから上述の特徴パラメータを使用して隠れマルコフモデルにより事前に作成しておく。5は音素照合部であり、特徴抽出部2の出力する音素の特徴量を音素モデル6の対応する音素の特徴量と照合して照合スコアを出力するところである。7は継続時間長照合部であり、特徴量である継続時間長を対応する継続時間長モデル8の特徴量により照合する。継続時間長照合部7が必要とする特徴量である継続時間長は、ネットワーク探索部3において累積照合スコアの計算時にネットワークの継続時間長単位の最終状態毎に得られる。音節単位の継続時間長であれば音節の最後の状態毎に得られる。

【0020】8は継続時間長モデルであり、継続時間長単位毎に、例えば音節毎に継続時間長の分布をラベル付けされた学習データから推定して事前に作成しておく。ネットワーク探索部3において音声認識時に使用するネットワークとしては、語彙を仮定しない場合は、音節単位にネットワークを形成した音節ネットワークを使用する。促音および撥音に関して日本語に存在する並び、例えば、促音は主に無声子音に接続するという様な条件を満足する様に接続する。語彙を仮定した場合は、語彙に依存したネットワークを作成し、これを使用する。また、ネットワーク中の各状態に隠れマルコフモデルの状態を、各音節の最終状態から次の音節への遷移に音節継続時間長の分布を、それぞれ対応させておく。

【0021】ここで、この発明の音声認識装置は、認識時には音声入力開始からネットワーク探索部3において累積照合スコアおよび必要な情報の伝播を開始する。音節ネットワーク上の前時刻(フレーム)までに累積照合スコアが伝播してきた各状態において累積照合スコアを計算する。この時、累積照合スコア計算部4へ照合要求を出し、対応する音素の照合を音素モデル6を使用して音素照合部5において照合する。即ち、特徴抽出部2に音声入力が開始されると、特徴抽出部2は入力される音声をデジタル信号に変換し、更に、これを1フレーム、例えば10ms毎に特徴パラメータに変換してその特徴量を音素照合部5に出力する。音素照合部5においては、特徴抽出部2の出力する音素の特徴量を音素モデル6の対応する音素の特徴量と照合し、照合スコアを累積照合スコア計算部4に出力する。累積照合スコア計算部4は音素照合部5から出力された照合スコアをネットワーク探索部3に送り込む。ネットワーク探索部3にお

いては、Viterbiアルゴリズムに基づいて、図1に示される様な幾つかの経路が交わる状態或は節点において累積照合スコアが最大になる経路のみを残す操作をする。ここで、節点においては、選択された経路を後でたどることができるものに、前の節点からの時刻とその節点に対する接続方向の情報を残す。これらの操作を音声終端に達するまでフレーム毎に繰り返し、音声終端においてはネットワーク上の最大累積照合スコアを持つ経路を選択して認識結果とする。

【0022】一方、音素照合部5における音素照合に対応して、継続時間長照合部7において特徴量である継続時間長を対応する継続時間長モデル8により照合し、採用した継続時間長単位からの遷移に対する計算をする。ところで、この特徴量である継続時間長は、ネットワーク探索部3において累積照合スコアの計算時にネットワークの継続時間長単位の最終状態毎に得られる。音節単位の継続時間長であれば音節の最後の状態毎に得られる。継続時間長を計測するために、図4により図示説明される様に、次の音節へ遷移する時刻を保持し、累積照合スコアとともに伝播させる。継続時間長単位の最終状態から遷移するときに、現時刻t₁と伝播してきた時刻t₂との差γ=t₁-t₂をとり、これが継続時間長となる。

【0023】ここで、継続時間長モデル8は、継続時間長単位毎に、例えば音節毎に継続時間長の分布をラベル付けされた学習データから推定して事前に作成しておいたものであり、認識時にはこの分布を確率密度関数として使用し、音節の最終状態から次の音節の初期状態へ遷移する時に音節継続時間長に対する確率を算出し、これを照合スコアとして累積照合スコア計算部4に返す。

【0024】音素照合部5および継続時間長照合部7それぞれの照合スコアを累積照合スコア計算部4に返し、ここにおいて照合スコアに重みづけその他の処理を施して累積照合スコアとした後、これをネットワーク探索部3に返す。以上の処理を入力音声の終端まで繰り返した後、ネットワークの終端において得られる最適なパスをたどり、これを認識結果とする。

【0025】この様に、累積照合スコアと状態遷移時刻とを組にして伝播することにより状態より長い系列を単位とする継続時間長制御をすることができる。この単位として音節を使用することにより、より頑健な継続時間長制御をすることができ、語彙の制約がない場合の挿入誤りを効果的に防止することができる。

実施例3

隠れマルコフモデルによる標準パターンの他に、情報としてそれまでの言語的履歴、例えば、現状態に到達するまでの音節の履歴を使用し、確率的な言語モデル、例えば音節連鎖確率により言語的制約をする場合について説明する。実施例1における継続時間長制御のための遷移時刻の伝播と同様に、各状態毎に生じる経路に対して累

積合スコアと共に履歴を保持し、伝播する。言語的な単位、例えば音節の終端において言語モデルによる尤度を計算して累積照合スコアに反映した後、言語的履歴を更新し、伝播する。これにより隠れマルコフモデルと確率的言語モデルを同時に使用することができる。

【0026】実施例4

隠れマルコフモデルによる標準パターンの他に、情報としてそれまでの継続時間長の履歴を使用し、現在の継続時間長を予測して継続時間長制御をする場合について説明する。実施例1における継続時間長制御のための時刻の伝播と同様に各状態毎に生じる経路に対して累積照合スコアと共に継続時間長の履歴を保持してこれを伝播する。継続時間長単位、例えば音節の終端において継続時間長の計算と尤度を計算し、累積照合スコアに反映した後、継続時間長の履歴を更新し、伝播する。これにより、隠れマルコフモデルと予測的な継続時間長制御をすることができる。

【0027】以上の通り、この発明は、隠れマルコフモデルによるネットワークを状態と節点により表現し、このネットワーク上においてViterbiアルゴリズムにより各状態に生じる音声認識候補について認識処理に必要な項目をすべて累積照合スコアと組にして伝播、処理することにより、累積照合スコアの計算量を減らし、記憶量も比較的小さく、隠れマルコフモデルにまたがった継続時間長制御をすることができる音声認識方法を構成することができる。

【0028】そして、音声認識時に各状態に生じる累積照合スコアそれぞれに継続時間長の単位により状態遷移が起きた時の時刻を保持しておき、これを累積照合ス

コアと組にして伝播させて行くことにより隠れマルコフモデルにまたがった継続時間長制御をすることができる。次の継続時間長の単位からの遷移の時に、その単位に留まった継続時間長を計算し、事前に用意した継続時間分布に対する尤度を得てこれを累積照合スコアに含めることにより、継続時間長制御をすることができる。これにより、日本語は音節単位でほぼ一定のリズムを持つという特性を利用し、より頑健性のある音節単位の継続時間長制御を行なって認識精度を改善することができる。同様に、言語の連鎖確率或は継続時間長を予測的に使用する場合においても利用できる。

【0029】

【発明の効果】以上通りであって、この発明は、隠れマルコフモデルのネットワークを状態と節点により表現し、累積照合スコアと共に認識処理に必要な他の項目を伝播させることにより、比較的小さい記憶容量で認識装置を構成することができると共に、累積照合スコアを効率よく計算することができ、更に隠れマルコフモデルを構成する状態の単位より長い状態系列の制御をする音声認識方法を構成することができる。

【図面の簡単な説明】

【図1】ネットワークの実施例を説明する図。

【図2】隠れマルコフモデルの一部の定義を説明する図。

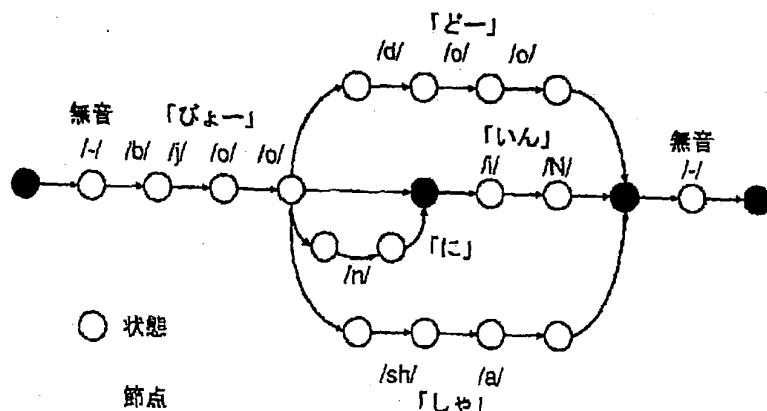
【図3】状態、節点を使用するネットワークの効率的計算の仕方を説明図。

【図4】継続時間長制御を用いた場合の処理方法の概念を示す図。

【図5】音声認識装置実施例を示すブロック図。

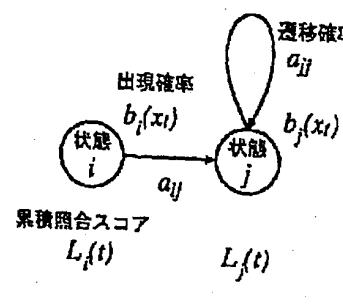
【図1】

図1



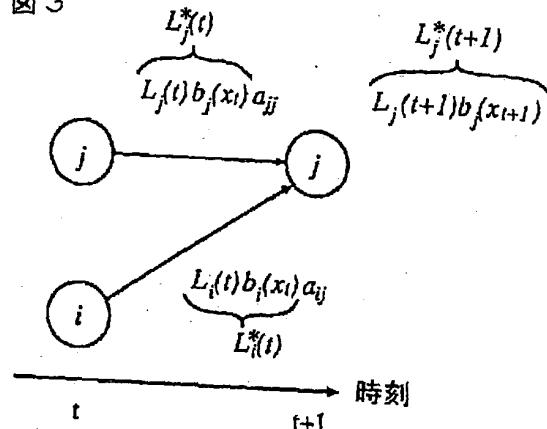
【図2】

図2



隠れマルコフモデルの定義

図3



【図4】

図4

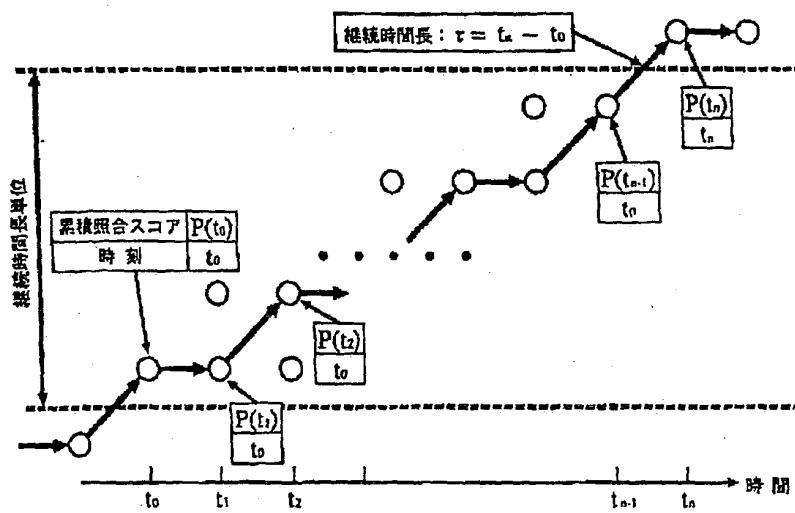


図5】

